

PROCESO ESTADISTICO



1. Planteamiento del problema

Toda investigación comienza con una pregunta, por ejemplo:

¿Es perjudicial beber demasiada agua?

¿Cuál es el coste de la vida en San Francisco?

¿Quién ganará las próximas elecciones presidenciales?

¿Los productos de herbolario realmente ayudan a tener buena salud?

¿Renovarán mi serie de televisión favorita el año que viene?

Ninguna de estas preguntas habla directamente de cifras. Sin embargo, todas ellas requieren el uso de datos y procesos estadísticos para llegar a una respuesta.

Imagina que un investigador quiere saber quién ganará las próximas elecciones presidenciales Colombia. Para responder con conocimiento de causa, el investigador debe seguir varios pasos:

Este es el punto de partida de toda investigación y define el eje central del estudio. La pregunta debe ser formulada de manera que pueda ser respondida con datos de muestra.

o Interrelación: Una pregunta clara y precisa guía todo el proceso subsiguiente, desde la selección de datos hasta el tipo de análisis y las conclusiones.

2. Recogida de datos

Determinar la población objeto de estudio.

Ejemplo el investigador pretende estudiar a los ciudadanos mayores de edad que tengan previsto votar en las siguientes elecciones.

Recopilar los datos.

- o Consiste en obtener la información de interés, que puede ser interna o externa (publicada o de primera mano). Es fundamental que la información se considere "dato estadístico", es decir, que sus valores permitan analizar relaciones significativas, siendo capaces de ser comparados, analizados e interpretados.
- o La calidad de los datos es crucial; deben ser fiables, no sesgados y exactos. Los métodos de recolección deben planificarse cuidadosamente para evitar sesgos, que son "favoritismos sistemáticos" hacia determinadas personas o respuestas.
- o Interrelación: La calidad y representatividad de los datos recopilados son la base para la validez de cualquier análisis posterior y las conclusiones..



3. Organizar, resumir y analizar los datos.

Una vez el investigador ha recopilado todos los datos que necesita, organizarlos, resumirlos y analizarlos le ayuda a responder la pregunta planteada. Este paso es el que la mayoría de la gente asocia con la estadística.

- o Los datos recopilados deben ser ordenados, corregidos si es necesario, y clasificados adecuadamente para su tabulación y para que sean legibles y procesables. Esto implica resumir, procesar e integrar la información.
- o Las tablas de frecuencias son un método común para resumir datos, agrupándolos por intervalos o clases cuando hay un gran número de valores.
- o Interrelación: La organización permite que los datos brutos sean comprensibles y preparan el camino para su presentación y análisis.

4. Análisis de los datos: Examinar todos los resúmenes de datos, gráficos, diagramas y análisis y extraer conclusiones de ellos para intentar responder la pregunta original.

- o Aquí es donde los diferentes tipos de análisis estadístico se interrelacionan para profundizar en la comprensión de los datos y responder a la pregunta de investigación.
- o Estadística Descriptiva Numérica: Consiste en calcular medidas que "describen" cuantitativamente los datos y resumen la información.

- Medidas de Tendencia Central: Indican la localización del centro de la distribución. Las más utilizadas son la media aritmética (promedio, sensible a valores extremos), la mediana (valor central, preferible con valores extremos o distribuciones asimétricas) y la moda (valor más frecuente, aplicable a datos categóricos).
- Medidas de Dispersión: Cuantifican la variación o el agrupamiento de los datos respecto al centro. Incluyen el rango (diferencia entre máximo y mínimo), la varianza y la desviación estándar (mide la dispersión promedio en las mismas unidades que los datos) y el coeficiente de variación (medida relativa para comparar dispersión entre conjuntos con diferentes unidades o medias). Estas medidas complementan la media, indicando el "mayor o menor agrupamiento de los datos". La regla empírica (68-95-99,7) relaciona la media y la desviación estándar en distribuciones normales para describir el rango de la mayoría de los datos.
- Medidas de Posición No Centradas: Cuantiles como cuartiles, deciles y percentiles, que dividen los datos ordenados en partes iguales. La puntuación Z estandariza una observación para comparar su posición relativa en diferentes distribuciones.
- Medidas de Forma: Como la asimetría (sesgo) y la curtosis (apuntamiento), que describen la forma de la distribución.
- o Análisis de Relación: Examina las conexiones entre variables.
- La correlación cuantifica la fuerza y dirección de una relación lineal entre dos variables numéricas.

5. Presentación de la información: Informes claros, impecablemente presentados con información de la metodología empleada y explicaciones simples pero precisas de los resultados obtenidos

- o Los datos organizados se pueden presentar de tres maneras principales: mediante enunciados o textos (para pocos datos), tablas estadísticas (para resumir gran cantidad de datos y facilitar comparaciones), y gráficas estadísticas.
- o La representación gráfica es crucial porque ofrece una visión de conjunto del fenómeno investigado más rápidamente perceptible y fácil de recordar que los números. La elección del gráfico depende del tipo de variable (cualitativa o cuantitativa) y las características a destacar.
- Para variables cualitativas: Gráficos de barras y de sectores (circulares). Una variante es el diagrama de Pareto, que ordena las categorías por frecuencia para comparaciones.
- Para variables cuantitativas: Histogramas, polígonos de frecuencias, gráficos de tallo y hojas (stem-and-leaf plots), gráficos de puntos (puntiogramas) y diagramas de caja y extensiones (boxplots). Estos visualizan la forma, el centro y la dispersión de los datos. El boxplot, por ejemplo, resume cinco estadísticos clave: mínimo, primer cuartil, mediana, tercer cuartil y máximo, y ayuda a identificar valores atípicos, simetría o asimetría, y a comparar muestras.
- o Los gráficos también ayudan a identificar el patrón de comportamiento de una distribución, como la presencia de modos (unimodal, bimodal, multimodal) o la simetría/asimetría.
- o Interrelación: La presentación visual de los datos es una "primera incursión en el análisis", revelando patrones que luego serán cuantificados por el análisis numérico. Sin embargo, los gráficos tienen limitaciones y no siempre son suficientes para una descripción profunda.

TALLER # 2

Fases del Proceso Estadístico

Objetivo: Identificar y aplicar las etapas de planteamiento del problema, recolección y organización de datos.

Actividad 1: Planteamiento del problema

Define un problema estadístico y formula una pregunta de investigación que pueda responderse con datos.

Justificación: Explica por qué es importante responder esta pregunta y qué tipo de datos necesitarías.

Actividad 2: Recolección de datos

Población y muestra: Define la población objetivo y propone una muestra representativa para tu pregunta.

Actividad 3:

Diseño de instrumento:

Fuente Primaria: Diseña una breve encuesta de 3 preguntas para recolectar los datos necesarios.

Fuente secundaria: Busca en internet 1 dirección o una entidad donde puedas obtener información del tema planteado.